

Australasian Language Technology Association Workshop 2012

(ALTA 2012)

**Dunedin, New Zealand
4 – 6 December 2012**

Editors:

**Paul Cook
Scott Nowson**

**ISBN: 978-1-62748-403-9
ISSN: 1834-7037**

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2012) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2013)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

Contents

Invited talks	1
<i>Using a large annotated historical corpus to study word-specific effects in sound change</i> Jennifer Hay	2
<i>Diverse Words, Shared Meanings: Statistical Machine Translation for Paraphrase, Grounding, and Intent</i> Chris Brockett	3
Full papers	4
<i>A Citation Centric Annotation Scheme for Scientific Articles</i> Angrosh M.A., Stephen Cranefield and Nigel Stanger	5
<i>Semantic Judgement of Medical Concepts: Combining Syntagmatic and Paradigmatic Information with the Tensor Encoding Model</i> Michael Symonds, Guido Zuccon, Bevan Koopman, Peter Bruza and Anthony Nguyen	15
<i>Active Learning and the Irish Treebank</i> Teresa Lynn, Jennifer Foster, Mark Dras and Elaine Uí Dhonnchadha	23
<i>Unsupervised Estimation of Word Usage Similarity</i> Marco Lui, Timothy Baldwin and Diana McCarthy	33
<i>Valence Shifting: Is It A Valid Task?</i> Mary Gardiner and Mark Dras	42
<i>Experimental Evaluation of a Lexicon- and Corpus-based Ensemble for Multi-way Sentiment Analysis</i> Minh Duc Cao and Ingrid Zukerman	52
<i>Segmentation and Translation of Japanese Multi-word Loanwords</i> James Breen, Timothy Baldwin and Francis Bond	61
<i>Measurement of Progress in Machine Translation</i> Yvette Graham, Timothy Baldwin, Aaron Harwood, Alistair Moffat and Justin Zobel	70
<i>Towards Two-step Multi-document Summarisation for Evidence Based Medicine: A Quantitative Analysis</i> Abeed Sarker, Diego Mollá-Aliod and Cecile Paris	79
<i>In Your Eyes: Identifying Clichés in Song Lyrics</i> Alex G. Smith, Christopher X. S. Zee and Alexandra L. Uitdenbogerd	88
<i>Free-text input vs menu selection: exploring the difference with a tutorial dialogue system.</i> Jenny McDonald, Alistair Knott and Richard Zeng	97
Short papers	106

<i>langid.py for better language modelling</i>	Paul Cook and Marco Lui	107
<i>LaBB-CAT: an Annotation Store</i>	Robert Fromont and Jennifer Hay	113
<i>Classification of Study Region in Environmental Science Abstracts</i>	Jared Willett, Timothy Baldwin, David Martinez and Angus Webb	118
ALTA Shared Task papers		123
<i>Overview of the ALTA 2012 Shared Task</i>	Iman Amini, David Martinez and Diego Molla	124
<i>Automatic sentence classifier using sentence ordering features for Event Based Medicine: Shared task system description</i>	Spandana Gella and Duong Thanh Long	130
<i>Feature Stacking for Sentence Classification in Evidence-Based Medicine</i>	Marco Lui	134
<i>Experiments with Clustering-based Features for Sentence Classification in Medical Publications: Macquarie Test's participation in the ALTA 2012 shared task.</i>	Diego Mollá	139